

AD-A185 952

AUTOMATION OF THE DUPLICATE CHECKING FUNCTION AT DTIC  
(DEFENSE TECHNICAL (U) DEFENSE TECHNICAL INFORMATION  
CENTER ALEXANDRIA VA OFFICE OF I... A D KUHN ET AL.  
MAY 82 DTIC/TR-87/21

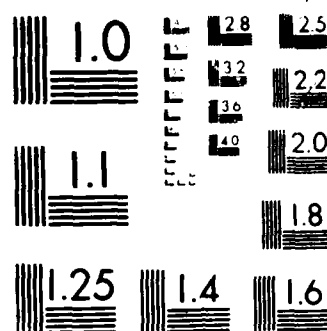
1/1

UNCLASSIFIED

F/G 5/2

NL





2. RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A



# **AUTOMATION OF THE DUPLICATE CHECKING FUNCTION AT DTIC**

## **Part 1: Overview**

Allan D. Kuhn  
Gladys A. Cotter

May 1982

**AD-A185 952**

**D  
T  
I  
C**

**Defense  
Technical  
Information  
Center**

**Office of Information Systems and Technology**

**Cameron Station, Alexandria, VA 22304-6145**

Unclassified  
SECURITY CLASSIFICATION

AD-A185 952

DOCUMENTATION PAGE

1a. REPORT SECTION Unclassified			1b. RESTRICTIVE MARKINGS <b>DTIC FILE COPY</b>						
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.						
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			5. MONITORING ORGANIZATION REPORT NUMBER(S)						
4. PERFORMING ORGANIZATION REPORT NUMBER(S) DTIC/TR-87/21			7a. NAME OF MONITORING ORGANIZATION <b>DTIC</b>						
6a. NAME OF PERFORMING ORGANIZATION Defense Technical Information Center		6b. OFFICE SYMBOL (if applicable) DTIC	7b. ADDRESS (City, State, and ZIP Code) <b>DTIC</b> <b>NOV 23 1987</b>						
6c. ADDRESS (City, State, and ZIP Code) Cameron Station Alexandria, VA 22304-6145		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER <b>A</b>							
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (if applicable)	10. SOURCE OF FUNDING NUMBERS						
8c. ADDRESS (City, State, and ZIP Code)		<table border="1"><tr><td>PROGRAM ELEMENT NO. 65801S</td><td>PROJECT NO.</td><td>TASK NO.</td><td>WORK UNIT ACCESSION NO.</td></tr></table>				PROGRAM ELEMENT NO. 65801S	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
PROGRAM ELEMENT NO. 65801S	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.						
11. TITLE (Include Security Classification) Automation of the Duplicate Checking Function at DTIC. Part 1: Overview.									
12. PERSONAL AUTHOR(S) Allan D. Kuhn and Gladys A. Cotter									
13a. TYPE OF REPORT		13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) 820500	15. PAGE COUNT 24					
16. SUPPLEMENTARY NOTATION									
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)						
FIELD	GROUP	SUB-GROUP	Automation, Duplicate checking, Card catalogs, Defense Technical Information Center, DTIC						
5	2								
12	5								
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Prior to 1981 the Defense Technical Information Center (DTIC) used manual processes inhouse for technical report collection, acquisition, duplicate checking, cataloging, and reference procedures. These procedures were accomplished through DTIC's card catalog, which referenced a little over one million technical reports, and hypothetically paralleled the same records stored in the DTIC Technical Report Database. The card catalog's primary justification was to serve as a backup to the database. But in practice it was the principal source of inhouse reference to the Technical Report collection. This report explains the reasons for automating duplicate checking to make online use of the database, and shows how the manual processes both affected the effort and were effected by it.									
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified/Unlimited						
22a. NAME OF RESPONSIBLE INDIVIDUAL Gladys A. Cotter			22b. TELEPHONE (Include Area Code) (202) 274-5367	22c. OFFICE SYMBOL DTIC-EB					

[illegible]

May 1982

Dist	Special
111	

[illegible]

## INTRODUCTION

Prior to 1981 the Defense Technical Information Center (DTIC) used manual processes inhouse for technical report selection, acquisition, duplicate checking, cataloging, and reference procedures. This entire set of manual procedures was done through DTIC's card catalog, which referenced a little over 1,000,000 technical reports, and hypothetically, paralleled the same records stored in the DTIC technical report, computerized, online data base. The card catalog's primary justification was to serve as a backup to the data base. But in practice it was the principal source of inhouse reference to the technical report collection.

There were two exceptions to manual procedures making use of the card catalog, one inhouse and the other outside. Inhouse, an office of professional level technical information specialists had been online several years, servicing bibliographic subject search requests for DTIC users. The other exception, much more important for the purpose of this paper, was the DTIC users themselves. Those who had direct access to DTIC's Defense RDT&E On-Line System (DROLS) had only the online data base available to them, with no access at all to the card catalog.

This lack of access to the card catalog is important for the reason that the catalog information was more current than the storage in the data base by four to six weeks. The DROLS users, therefore, lacked displayable confirmational data by that amount of time, meaning in essence, that a technical report record could be in process in the system, but the online users, including those inhouse, would not have retrieval and order access to a document for that length of time. Online access to technical reports, then, was delayed by the manual procedures practiced at that time.

In order to increase access it was necessary to solve the problems of manual procedures through automation solutions. It was also necessary to justify the benefits of automation by recognizing and stating the requirements for automation, both organizational and procedural. I will show, therefore, how automation was accomplished, centered around the technical report duplicate checking function, which was the first procedural point of card catalog use for technical report processing.

#### TO 1981: DTIC TECHNICAL REPORT INPUT MANUAL PROCEDURES

Under the older system, a mix of manual processing and data base input processing was employed. But the data base input was dependent on the accomplishment of the manual procedures, which is why the accessibility of data base information was delayed, explained as follows.

#### The Computer Input Process

DTIC uses a UNIVAC 1100/80 mainframe computer system for its DROLS information handling. There are several data bases resident in the system, storing information on planned, ongoing, completed, and independent DoD related research and development (R&D). This paper deals with the completed R&D, Technical Report (TR) Data Base because this data base system was built around the initial purpose of DTIC's predecessor organization,\* collecting, controlling, and disseminating DoD-generated R&D technical reports, with all

---

\* Defense Documentation Center.

processes and controls incorporated therein for doing so.

The TR data base is basically divided into two parts for handling technical report records, the Main Accessioned Document (MAD) file and the Current File (CF). The MAD file, on disk drive auxilliary storage, is the principal storage of the Accessioned Document (AD) collection, with a little over one million records. The Current File, utilizing magnetic tape and temporary disk storage through various processing phases, is a holding file of newly-accessioned documents currently being inputted into the TR data base. There are three two-week cycles of ADs in the Current File, one in a state of ready for release to the MAD file, one in a state of being reviewed for quality control, and one in a state of currently being built. Each cycle averages 1200 AD records.

Under the older system, no Current File data was displayable, and had to wait for release to the MAD file before being so. The CF could be searched, but the lack of display capability, except for the AD record number, precluded confirmation of finds. The online users' only recourse to confirm a find was to correspond with and provide bibliographic information to the DTIC Reference staff. They in turn referred manually to the card catalog to confirm retention or recent entry, since the catalog was the only source of current bibliographic information, i.e., title, report number, source, date, and contract. If the user wanted merely a yes-or-no answer, the manual procedures, including the communication, greatly extended time by several days to a week in getting that information. If the user should want to order the document after confirmation, however, then the extended manual time was not all that critical since the report could not be ordered,



anyway, until the computer record was in the MAD file from where its order could be computer processed.

In the manual and computer input mix, DTIC had evolved from a centrally located Text Processing Subsystem (TPS) where the manually processed information was input by Data Transcriber technicians through paper tape punching and reading, to actual online input through a Remote Terminal Input Subsystem (RTIS). The RTIS hardware i/o devices were UNIVAC 200 CRT terminals, hardwired into the mainframe computer system.

In a sense, the installation of these i/o terminals at several inhouse stations constituted a rudimentary but incipient input/output network, which will be referred to shortly. These terminals were still operated, however, by the Data Transcribers. They only keyed in information that had already been manually processed through duplicate checking, descriptive cataloging, subject analysis, and quality control.

#### The Manual Input Processes

The technical report bibliographic data, then, was never input until the manual processes were accomplished. The keying-in of the data was basically an activity redundant of the cataloging activities, but not necessarily of the duplicate-checking activity. A summary of the manual processes is as follows:

A newly arrived report, regardless of its date, was checked against the card catalog to determine whether it had already been accessioned or not, and by extension, whether or not it was already recorded in the data base. All current information for an accepted report was hand-transcribed on 3x5 cards which were interfiled in the card catalog, to await computer-

generated cards to arrive four to six weeks later after a CF cycle had been released.

Additionally, any acquisitions and selection information was also incorporated in the card catalog. This information ranged from documents on order to reports not available for accessioning, documents not falling within the selection standards for accessioning, and documents annotated for some reason or other as being in route, or returned, but expected to be accessioned. After the manual duplicate checking process and acceptance for processing, the document was sent its way through the cataloging processes, manually prepared for keying-in, as mentioned.

An additional factor in retaining the manual processes up until 1981 was the cost benefit value of such procedures versus the costs of obtaining the hardware for automation of the processes. With no outside influences to create the change from manual to automated, it was thought more costly to make the change yet basically retain the same functions. It was thought the same number of duplicate-checkers, catalogers, and the rest of the personnel along the pipeline functions would remain in place, the only change being the additional i/o equipment and programming for automated functions.

Studies had been made for eventual automation of all technical report processing functions. The costs of total automation, however, in terms of hardware needed as required with the redesign of the system, were not yet justifiable. Not yet taken into consideration was a change in the mission of the organization and developments in information transfer technology needs.

### Mission Change: Resource Sharing

DTIC's former mission as DDC was basically collecting DoD R&D technical reports and bringing them under computer control for dissemination in accord with DoD imposed requirements. Consequently, DDC and its TR data base system employed functionally centralized data handling techniques. All report input was done inhouse. DROLS allowed outside sites to retrieve and order, but only on the data base as created inhouse by DDC/DTIC.

When DDC became DTIC in 1979 its mission was changed to include the concept of not only storing records for DTIC computer-controlled documents for dissemination, but also records for DoD documents not at DTIC and with information on their place of availability. This concept was incorporated under a resource sharing experiment that had already been taking place since 1977, called the Shared Bibliographic Input Experiment (SBIE). This outside multisite technical report input experiment was done over a period of four years with the cooperation of six organization technical libraries.\* Five additional sites brought in to expand network input in the experiment in 1980.

The period throughout the time of the experiment was one of both successes and failures. Successes came in the manner of cooperative attitudes for technical report input both on the part of DTIC and of the sites. The cooperative interaction in turn resulted in developments of input standards both for network input and for easier retrieval of that

---

\* Naval Research Lab., AF Weapons Lab., Defense Nuclear Agency, Defence Communications Agency, Army Research & Development Command, and Institute for Defense Analyses.

standardized data.

Failures occurred, however, as the experiment slowly attempted a true shared cataloging input, i.e., open input of all documents by the sites. SBIE was set up to progress through two distinct phases. The first was input of an site's own organization reports, only, and DTIC not entering them. The second phase, however, was to be all documents received by and in the collection of the technical library site. This meant that a site would input any document immediately on receiving it on primary distribution, as would also DTIC.

At this juncture, the failure occurred with the DTIC MAD file, with its lacking of the current four to six week old bibliographic information and with that data base only available online for duplicate checking to the remote sites. At the same time DTIC was using its up-to-date manual card catalog, and in not using the online system for duplicate checking, never knew whether or not an SBIE site had entered a record. Both DTIC and the sites began duplicating each other's input in this phase. Some sites dropped the phase, feeling that record duplication was cost-deficient. The phase was halted with mutual agreement of both DTIC and the cooperating sites when in the case of one site the duplicate rate on their input of 100 records reached 50% in mid 1980.

This operational failure showed that DTIC's TR data base input system, in having been basically established for centralized computer control of technical report dissemination, was rigidly incompatible with the requirements of the network input and information processing concept.

The problem that caused this failure was defined as follows:

1. There was a disparity in duplicate checking functions.
2. The SBIE sites had available to them the TR data base, only, for

duplicate-checking, with displayable information only in the MAD file; that data was behind by four to six weeks, which was no good for current reports received on primary distribution.

3. DTIC procedures prescribed manual duplicate checking and entry of current information in the card catalog, with no use made of the online system whatsoever for duplicate checking.

It was the DTIC duplicate checking function, therefore, that became the pivotal point for deciding whether or not the SBIE program would work. In that the goal of SBIE was to become a DoD-wide online catalog of DoD technical reports and their availability, it was felt that automation of the function was both necessary and justifiable in order to create a uniform, common, up-to-date point of reference for duplicate checking for all inputting sites in the network, meaning in particular, of course, DTIC. The failure of shared input, therefore, generated the appropriate requirement for justifying online duplicate checking inhouse. The automation of this function had a ripple effect of further automation, which will be explained further.

#### THE PROGRAM FOR AUTOMATION

A responsive online duplicate checking capability was the definition given to the first requirement for a successful Shared Bibliographic Input network. The technical report Current File was redesigned to achieve this

goal. The redesign effort consisted of:

1. Daily extract of all RTIS records stored in all RTIS activities, inhouse and outside.
2. Daily processing of selected data fields from these records into a displayable Current Direct File that contains:

- field 1 - AD number
- field 5 - Corporate author
- field 6 - Unclassified title
- field 8 - Title classification (U)
- field 9 - Descriptive note
- field 10 - Personal authors
- field 11 - Report date
- field 14 - Report numbers
- field 15 - Contract numbers
- field 18 - Monitor acronyms
- field 19 - Monitor series numbers
- field 20 - Report classification
- field 34 - Report serial
- field 35 - Source code

3. Daily creation of a searchable Current Inverted File that contains search terms from the following Current Direct File data fields:

- field 6 - Unclassified title key (role\* 55 or 56)
- field 10 - Personal authors (role 11)
- field 11 - Report date (role 24)
- field 14 - Report numbers (role 51)
- field 15 - Contract numbers (role 16)
- field 18 - Monitor acronyms (role 03)
- field 19 - Monitor series numbers (role 53)
- field 20 - Report classification (role 58)
- field 34 - Report serial (role 52)
- field 35 - Source code (role 02)

This new current file made it possible for DTIC to store preliminary cataloging information online and proceed to automate its duplicate checking procedures. The following summary details the program for DTIC's automation of duplicate checking from January 1981, to March 1981 when the system was declared operational.

---

\* System search demand codes in the inverted file.

On 15 January 1981, the duplicate checking personnel began the online duplicate checking experiment. Before this date the duplicate checkers had been given training in the areas applicable to online duplicate checking: cataloging, retrieval, and input. In that the dup-checkers were to input skeletal cataloging data, they in effect became preliminary catalogers and were designated as such. Three UNIVAC 200 i/o terminals were implaced for use by the preliminary catalogers.

Online duplicate checking, combined with preliminary cataloging input, was carried out in three phases. The purpose of the phased implementation was to allow the preliminary catalogers to become experienced, procedures to be modified, and additional equipment needs to be determined and met.

The first phase covered technical reports that were typically received by the Shared Bibliographic Input Experiment (SBIE) sites. This phase excluded the online input of certain materials entered by DTIC and unlikely to be input by the SBIE sites, such as Patents, Patent applications, translations, and foreign documents. The second phase added foreign documents, and the third covered all document types. The only reports excepted in this experiment were those from SBIE sites or their contractors. These had already been dup-checked and input, online, by those sites.

During the initial period of the experiment the procedures for dup-checking were established as follows:

1. Dup-check the documents online using the title key role code 55 or 56.
2. When there is a find, display the record to verify that an actual duplication exists, or an antecedent record. If the document being dup-checked is a duplicate, send to document storage, or dispose of as appropriate. If the find displays as an antecedent record, follow the procedure indicating that only minimal processing is needed.

3. Those documents determined not to be duplicates in the online data base are to be rechecked in the card catalog for verification of the online results. If the card file check identifies a duplicate, record the AD number and the data element by which the duplicate was found. This information is to be included on a statistical form provided.

4. For the documents which pass the online and card file check, input the following data elements as a skeletal record:

Field 1 AD number

6 Title (unclassified)

8 Title classification (u)

11 Report date

14 Report number

15 Contract

18 Monitor acronym

19 Monitor report number

20 Report classification

5. Print three copies of the skeletal record. Insert the first copy in the document for descriptive cataloging use, insert the second copy in the main card catalog, retain the third copy for data acceptability review.

To track the experiment during this initial period, records for developing statistics were kept from 15 January to 13 February. The following table reflects the online production figures for this period:



# ONLINE PRODUCTION FIGURES, 15 Jan-13 Feb 1981

Date	Total Docs Sent Daily	Online Input	Online Dup-Chk	Comments
Jan 15	140	14 -----	same	Period of familiarization
16	84	12 -----	"	
19	140	30 -----	"	Online dup-checking approximated input
20	140	25 -----	"	
21	140	20 -----	"	
22	140	23 -----	"	
23	196	52 -----	"	
26	Ø*	17 -----	"	
27	140	19 -----	"	Online production increase period
28	140	25	15	
29	140	64	35	
30	140	50	32	
Feb 2	100	22	52	
3	170	34	28	Input includes documents dup-checked manually until online dup-checking became comparable to input.
4	150	24	71	
5	140	63	24	
6	Ø*	48	53	
9	140	60	28	
10	140	151	66	
11	140	68	0	
12	140	96	61	
13	140	150	109	
TOTALS	2800	1062	786	---approximate count; see comment on approximate input, above.

\* End of CF cycle

These data indicated that more documents were input online than dup-checked online. There were two factors influencing this situation. One factor was system down-time, during which documents were dup-checked manually in the card catalog. When the system was up again, skeletal record input was done for those documents. The second factor was that the two dial-up terminals, which were to be incorporated into the program, and had been ordered and planned for retrieval use in the experiment, were not received until the 16th of February.

During this same time period, records were kept on the number of duplicates located in the card file after an online search indicated that they were new to the system. Twenty duplicates were located in this way. The reasons they were not identified as duplicates during the online search were attributed to three basic causes:

1. Title variance; alteration of the title as it appears on the document, following DTIC cataloging policy at that time, when the document citation was entered into the TR data base.

2. Date variance; occurs with displays of different dates on the document and its document information form, which for DoD is DD 1473.

3. Input of the same document twice within a one-day time frame; the input duplication was caused by the overnight time lag of RTIS storage take-off into the Current File, creating a lag in dup-checking capability.

The problem of title variance was relevant to the DTIC acquisition and reference staffs as well as to the dup-checkers. The title key search capability requires use of the first five words of the title exactly as they were entered. This presents a problem for several reasons. One is that the acquisition and reference staffs are often called upon to locate reports with title information that is not as precise as the title key requires. For

instance, if words are transposed or substituted in the title information given, the title key does not work. Another problem is that the descriptive catalogers, in following established cataloging policy, are sometimes required to alter the title information as it appears on the cover. A simple example is:

Cover display: Quarterly Progress Report on Radar Tracking.  
Title entered: Radar Tracking.

A complex case of title entry variance is an example of an SBIE entry and DTIC entry on the same document. The SBIE title entry was JANNAF Combustion Meeting (17th), NASA Langley Research Center,... The DTIC entry was JANNAF Combustion Meeting (17th) Held at Hampton, Virginia,... The difference between the fifth words (NASA and Held) eliminated the usefulness of the title key search for dup-checking purposes.

In such cases, even with the document in hand, using the first five words of the title as they appear on the document does not generate a title search key which matches a duplicate record in the system. In order to solve the title variance problem, a request for a free text search capability for the title field was submitted. This capability is being reviewed.

In the case of date variance, a uniform policy of using the date on the cover was established. The justification for using the cover date is based on the instructions in the MIL-STD-847A\* format requirements for technical reports and the Form 1473, which requires that the date to be entered on the 1473 is that displayed on the cover. This policy was disseminated to the DTIC staff and the SBIE sites.

The input of the same document twice in one day is a problem that was anticipated as a built-in hazard of the system as it existed. A solution suggested by the SBIE sites in a strongly worded request was to develop an

---

\* The DoD standard for technical report format composition.

immediate update capability, similar to OCLC, which would solve this problem. The cost of programming this capability is still to be justified.

After a month of dup-checking one document at a time using the title key, role code 55 or 56, multiple document searching was initiated. Search strategy included titles of a group of documents either by the first five words of each title or by the algorithmic keys of the first five words, for example:

```
@scf@
?55adogshavtoma
?55tcowsjumovth
?55tspy*whocain
?55twoodboofam
end
```

Using this search strategy, documents were dup-checked in blocks. This procedure made dup-checking online more time-efficient than dup-checking in the card catalog.

At this point the preliminary catalogers were dup-checking and inputting information online in addition to maintaining the card catalog. In order to lessen this burden on the preliminary catalogers, it was necessary to identify uses of the card file which still were not satisfied by the online system. The needs of the reference, acquisition, preliminary cataloging, selection, and security specialist staffs were monitored. The following list is a compilation of the activities which were not accommodated by the online system, as well as any corrective actions planned or initiated.

1. The Supplementary Note, TR Field 21. This field was (and is) not searchable online because its information was considered merely supplemental. The field contains, however, overflow report numbers, contracts, references to cooperating organizations and joint efforts, and other valuable information. Under the manual procedures this information was entered and searchable in the card catalog. Our suggested Solution was a programming request for free text retrieval in this field. The cost of programming and inverting the data in this field was considered to be prohibitive.

2. Replaced (Superseded) and Cancelled Notes. These notes were not displayed online. When a record number that had been replaced or cancelled was displayed online, the only information available was the notice: "Not available for display." DTIC staff manually searched the card file to determine if an AD has been replaced or cancelled, and to determine the replacement number. Solution: A programming request for display of the notes, "This document cancelled; no longer available from DTIC," or "Replaced by AD-C123456," was submitted and accepted.

3. Security Reclassification/Distribution Change Notes. The notes on the authority received at DTIC to make changes in security/release status were transcribed only in the shelf list card file. Solution: A programming request for a new TR field to contain these notes was submitted and accepted. A benefit with this new item was that this particular information became available to online users.

4. Document Status Notations. Notations for documents which had been mailed to DTIC but were still in transit. These notes cover cases such as documents being recalled by the contributor after they have been sent to DTIC, or errata sheets arriving before the documents. These notes were placed in the card catalog to alert the dup-checkers to the status of the documents during the manual dup-check. Solution for automation: Skeletal input with a pre-assigned, reusable range of record numbers was set up. These skeletal records remain on the Current File until the document is received. This procedure was implemented during the experimental phase.

5. Acquisition and Selection Notations. These notations covered documents which were on order (blue cards), document requests that had been refused (green cards), and documents that had been returned after receipt as being out of scope for DTIC (yellow cards - for non-technical information, Official Use Only statements, illegible, etc.). The dup-checkers matched the documents they received with the blue cards in the card catalog. The yellow

and green cards alerted the DTIC staff to the reasons particular documents were not in the DTIC collection. This information was available only in the card catalog. Solution: A separate data base was created in DTIC's second mainframe computer, its older UNIVAC 1108, using BASIS DBMS to store this information. A program was written which matches the information on the 1108 against the Current File cycle and prints out matches between documents ordered and documents received. A match removes the record from the BASIS file, leaving the remaining records outstanding.

6. TIP's and ATI's. Information for these old predecessor organization reports from the 1940s and 1950s is available only in two static card catalogs. Solution: Continued referral use of these static card catalogs.

7. Older technical report records - "30 Year File." The 30-year file had been searchable but not displayable. DTIC staff referred to the card catalog when immediate information relating to those AD numbers was needed. Solution: Make the 30-year file displayable. Action on this matter had already been initiated prior to this experiment, and display became available effective as part of an overall effort to program display into the entire TR data base.

8. Report Numbers. Report numbers have been entered online in a standardized format in order to provide a structured access method into the data files, the DTIC TAB\* announcement indexes and DTIC special product indexes. To establish structured report numbers it has been necessary to apply rigid standardization concepts. This policy has succeeded for machine generation of indexes, but for search and retrieval applications it has never been practical. Attempts must be made by both inhouse and out-of-house users to ascertain how "standardized" numbers may have been entered. Report numbers can be located in the card file with little effort. Furthermore, a static card catalog requires that DTIC personnel outside of the Descriptive Cataloging

---

\* Technical Abstracts Bulletin, published biweekly.

Branch, where report number standards are recorded in a card file, spend time translating the report number as it appears on the cover into DTIC's standard in order to search the system or input report numbers. This problem was documented in more detail and attached to an earlier progress report submitted to DTIC. Solution: A programming request for a new TR field was submitted, to contain report numbers as they appear on documents when they are different from standardized display entries. The request was not implemented, the rationale being that displaying report number variants would be displaying conflictive information.

Besides the problem areas listed above, a quality unique to the card file was apparent -- it doesn't have down time! During the online dup-checking experiment the system was down or equipment was malfunctioning for some part of most days. But during the automation of duplicate checking and the acquisition and selection functions, it became acutely apparent that, with the necessity to automate in some manner all the card catalog procedures, the card catalog would no longer be needed, and its maintenance could be stopped.

It was determined, therefore, to make the card catalog static. After pro and con studies were made, a date for its closing was set at 1 Oct 81, which was adhered to. It was felt that leaving it open would only promulgate unnecessary card catalog dependence, in addition to not being cost justifiable.

A major argument against the closing of the catalog was that there would be no back-up for reference during system downtime. Back-up systems suggested included a COM file, video disk file, a tape cassette file, and continued maintenance of the card catalog shelf list. In spite of empirical complaints about large amounts of downtime, DTIC-Systems provided machine-generated statistics that showed that actual downtime was relatively very low. This effectively eliminated the cost justification for any kind of a back-up system.

## THE RIPPLE EFFECTS OF DUPLICATE CHECKING AUTOMATION

With automated duplicate checking becoming operational in Mar 81, and with the card catalog slated for closing in Oct 81, it was obvious that the descriptive cataloging function was a candidate for automation. The descriptive catalogers had extensively used the card catalog for referral purposes, and now would have to make use of data base retrieval to perform the same function. Direct data base searching required the need for terminals, which in turn presented the capability of online descriptive cataloging. A formal major project was consequently established, setting up the program for getting the catalogers directly online, vis-a-vis having their manual work keyed-in by Data Transcriber personnel. This functional automation took place after getting the i/o devices installed, beginning Oct 81,\* and was made operational in Feb 82.

During the study for descriptive cataloging automation it was further discerned that there was a choice of modular automated duplicate checking and descriptive cataloging functions, i.e., two separate operations, or that the functions could be streamlined by consolidating them. Consolidation was chosen because of its streamlining aspects. The two flow charts following indicate where consolidation streamlined the function. The arrows on the first chart, labeled separate cataloging functions, indicate the redundant processes. The second chart, labeled consolidated cataloging function, shows the reduced

---

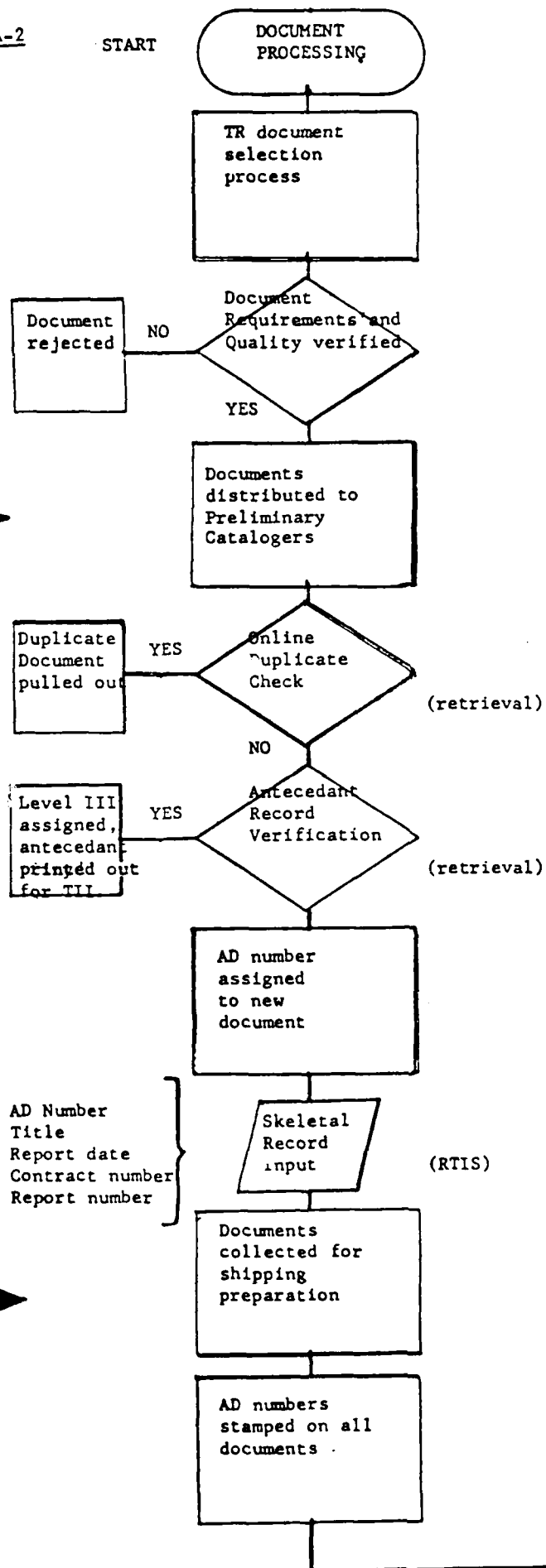
\* Original date for starting automation was Sep 81, but difficulties in equipment installation delayed that start date. The descriptive catalogers, who had already received their retrieval and input training, deftly made use of other i/o terminals for familiarization and search purposes when they saw the 1 Oct catalog closing date quickly approaching.



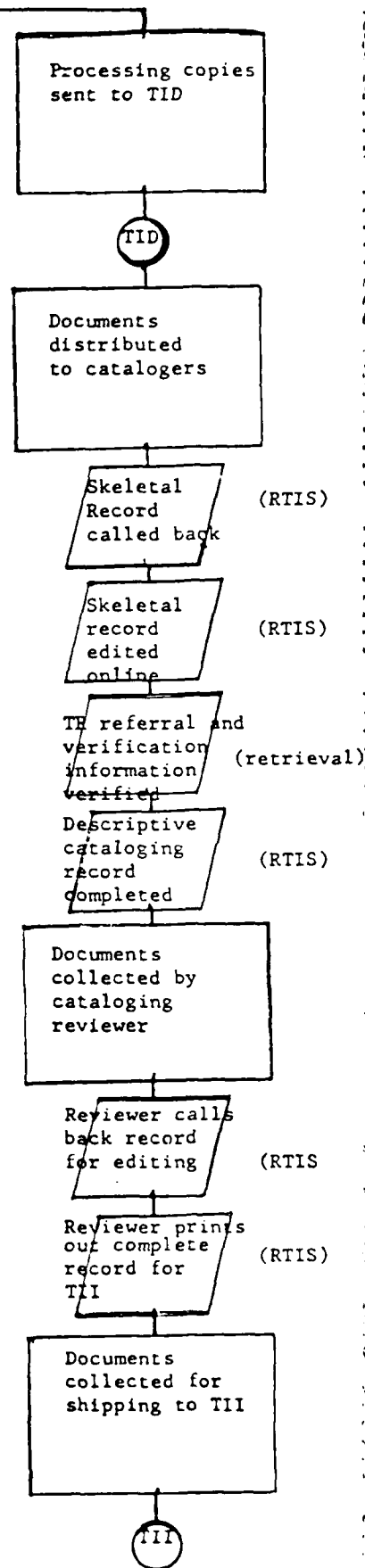
# SEPARATE CATALOGING FUNCTIONS

DDA-2

START



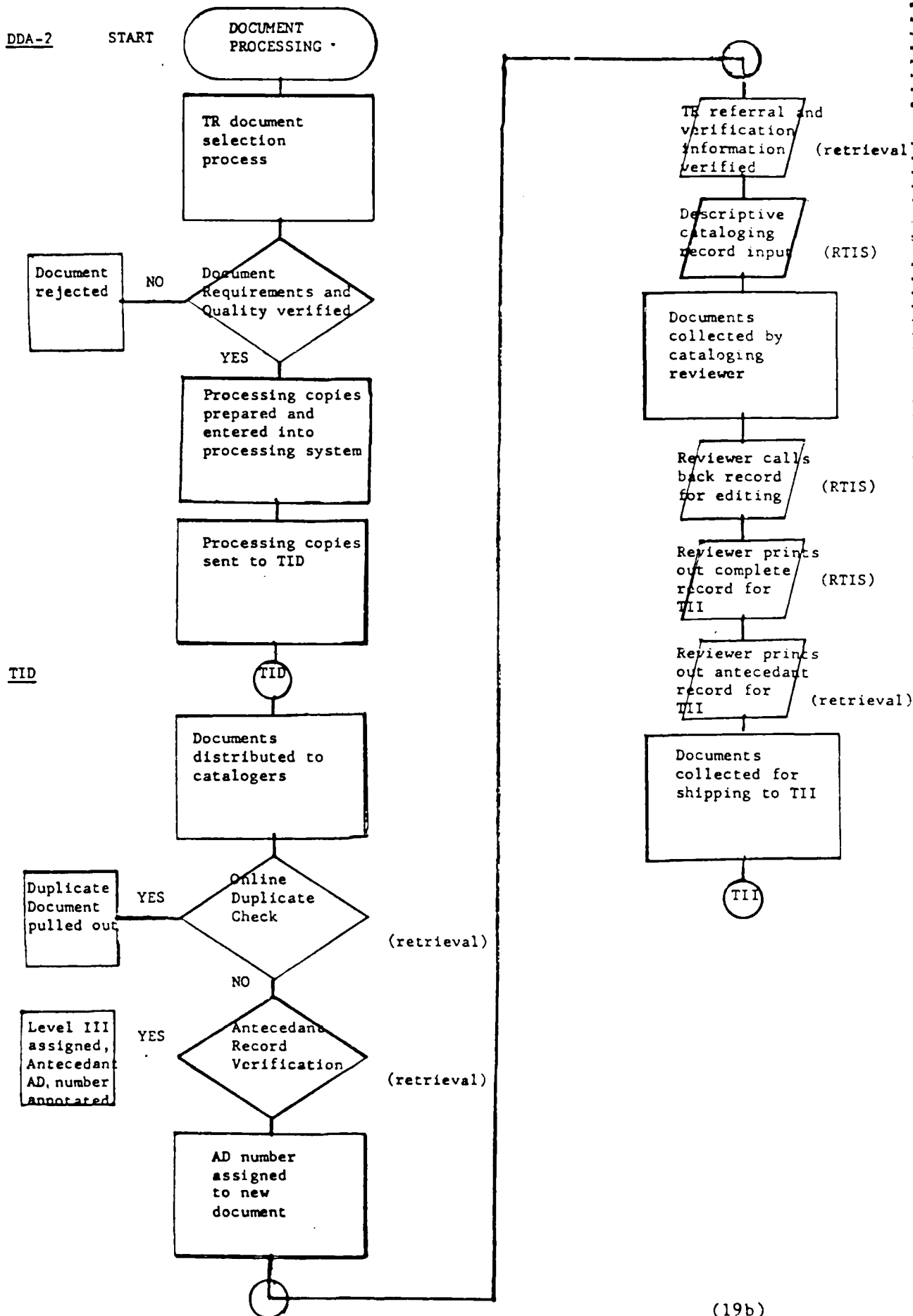
TID



# CONSOLIDATED CATALOGING FUNCTION

DDA-2

START



processing flow for both handling the documents and their online entry.

The consolidation eliminated the preliminary catalogers in that their function was placed in the Descriptive Cataloging Branch. The major concern, therefore, was the placement of these original duplicate checking personnel. OPM and Union policies over proper and appropriate placement rightfully prevented the consolidation for about 6 months until the personnel had all been placed. Functional consolidation finally did occur in Mar 82. Those original duplicate checkers, though no longer practicing their newly acquired skills, are to be considered the leading edge in this entire automation process at DTIC.

END  
DATE  
FILMED  
JAN  
1988